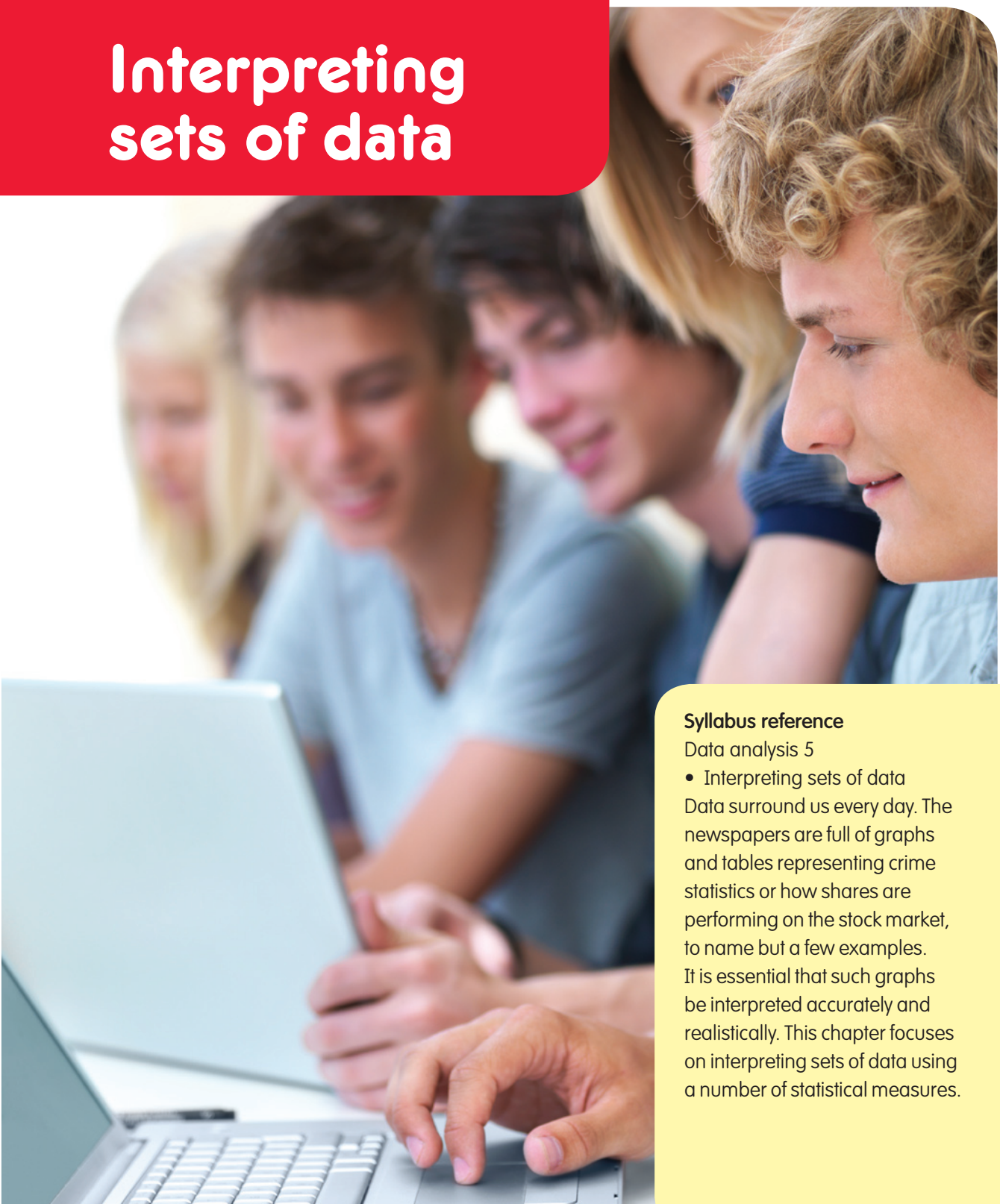
4

Interpreting sets of data



Syllabus reference

Data analysis 5

- Interpreting sets of data

Data surround us every day. The newspapers are full of graphs and tables representing crime statistics or how shares are performing on the stock market, to name but a few examples. It is essential that such graphs be interpreted accurately and realistically. This chapter focuses on interpreting sets of data using a number of statistical measures.

ARE YOU READY?

eBookplus
Digital doc
SkillSHEET 4.1
doc-1330
Finding the mean

eBookplus
Digital doc
SkillSHEET 4.2
doc-1331
Finding the mode

eBookplus
Digital doc
SkillSHEET 4.3
doc-1332
Finding the median

eBookplus
Digital doc
SkillSHEET 4.4
doc-1336
Finding the range

eBookplus
Digital doc
SkillSHEET 4.5
doc-1337
Finding the interquartile range

eBookplus
Digital doc
SkillSHEET 4.6
doc-1338
Choosing the appropriate standard deviation

eBookplus
Digital doc
SkillSHEET 4.7
doc-1341
Compiling a stem-and-leaf plot

eBookplus
Digital doc
SkillSHEET 4.9
doc-1349
Drawing a box-and-whisker plot

Try the questions below. If you have difficulty with any of them, extra help can be obtained by completing the matching SkillSHEET. Either click on the SkillSHEET icon next to the question on the eBookPLUS or ask your teacher for a copy.

Finding the mean

1 Find the mean of the following sets of scores.

a 3, 5, 8, 3, 9, 4, 3, 5

b

Stem	*Leaf*
0	9
1	2 2 3 6 7
2	4 5 7
3	0 0

c

Score	Frequency
1	9
2	14
3	32
4	27
5	18

Finding the mode

2 For each of the data sets in question **1** find the mode.

Finding the median

3 Find the median of the data sets in question **1**.

Finding the range

4 Find each of the data sets in question **1** find the range.

Finding the interquartile range

5 For each of the data sets in question **1** find the interquartile range.

Choosing the appropriate standard deviation

6 For each of the following choose and find the appropriate measure of the standard deviation.

a At the end of a movie 10 viewers were chosen and asked to rate the movie from 1 to 5 stars. The results were: 3, 4, 2, 3, 1, 5, 2, 4, 3, 5.

b At Yass High School there are 80 students who study General Mathematics. They all sat for a test scored out of 20, and the results obtained are given in the table below.

Score	14	15	16	17	18	19	20
Frequency	3	9	15	24	18	7	4

Compiling a stem-and-leaf plot

7 The scores below show the number of customers each day in a car yard.
23, 32, 27, 31, 19, 45, 22, 26, 38, 41, 27, 40, 9, 34, 37, 21,
22, 30, 39, 19, 14, 32, 20, 40, 23, 27, 26, 28, 11, 15, 28, 33
Display the data in a stem-and-leaf plot.

Drawing a box-and-whisker plot

8 For the data set in question **6**, display the results using a box-and-whisker plot.

4A Measures of location and spread

Consider the following set of scores that are the exam results for 10 students.

55, 57, 57, 58, 60, 60, 62, 63, 63, 65

To identify a score that is typical in this data set, we can use the mean or median.

- The *mean* is calculated by adding all the scores and dividing by the number of scores in the set.
  When the data is a small set of scores the mean is found using the formula

$$\bar{x} = \frac{\Sigma x}{n}$$

where $\bar{x}$ = mean
x = individual scores (Therefore, Σx represents the sum of individual scores.)
n = number of scores.

Where the data is presented in a frequency table we use the formula

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$

where $\bar{x}$ = mean
x = individual scores
f = frequency.

In this formula Σfx represents the sum of the frequency × score column on the frequency table and Σf represents the sum of frequency column.

- The *median* is the middle score (odd number of scores) or the average of the two middle scores (even number of scores).
  For this set of scores:

Mean = 600 ÷ 10
= 60
Median = 60

Both the mean and median are a measure of location within a data set.

WORKED EXAMPLE 1

For the set of scores 13, 19, 31, 40, 55, 65, 90, 92, 95, 100 calculate:
a the mean
b the median.

THINK	WRITE
a 1 Find the total of the scores.	a Total = 600
2 Divide the total by the number of scores.	Mean = 600 ÷ 10 = 60
b Average the two middle scores.	b Median = (55 + 65) ÷ 2 = 60

We have now examined two data sets. Look at these data sets side by side.

Set A: 55, 57, 57, 58, 60, 60, 62, 63, 63, 65
Set B: 13, 19, 31, 40, 55, 65, 90, 92, 95, 100

Although both sets of scores have the same mean and median, they are very different sets of scores. Clearly, in Set B the scores are more spread out than in Set A. To measure the spread of a set of scores, we use one or all of the following.

- Range: Highest score – lowest score

  Set A: Range = 65 – 55 = 10 Set B = 100 – 13 = 87

- Interquartile range (IQR): The difference between the upper quartile and lower quartile.

  Set A: Interquartile range = 63 – 57 = 6 Set B: Interquartile range = 92 – 31 = 61

- Standard deviation: Found using the σ_n (population) or s_n (sample) functions on the calculator.

  Set A: $\sigma_n = 3.07$ Set B: $\sigma_n = 31.51$

Each of these measures of spread show that in Set B the scores are more scattered than in Set A.

WORKED EXAMPLE 2

eBookplus Tutorial int-2417 Worked example 2

For the set of scores 45, 62, 75, 69, 50, 87, 92 calculate:
a the range **b the interquartile range** **c the standard deviation.**

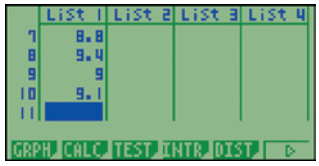
THINK	WRITE
a Subtract the lowest score from the highest score.	a Range = 92 – 45 = 47
b 1 Write the scores in ascending order.	b 45, 50, 62, 69, 75, 87, 92
2 Divide the data in two halves, leaving the middle score out of both sets.	45, 50, 62, 75, 87, 92
3 The lower quartile is the median of the lower half; the upper quartile is the median of the upper half.	Lower quartile = 50 Upper quartile = 87
4 Subtract the lower quartile from the upper quartile.	Interquartile range = 87 – 50 = 37
c Enter the set of scores into your calculator using the statistics function.	c $\sigma_n = 16.36$

WORKED EXAMPLE 3

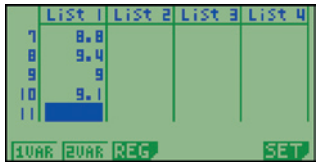
Nadia is a gymnast. For a routine she is given the following scores by 10 judges.
9.0 8.7 9.2 9.3 9.8 9.2 8.8 9.4 9.0 9.1
Use your graphics calculator to find
a the mean **b the median** **c the mode**
d the range **e the interquartile range** **f the population standard deviation**
g the sample standard deviation.

THINK	WRITE
1 From the **MENU** select **STAT**.	

2 Delete any existing data, and enter the scores above in **List 1**.

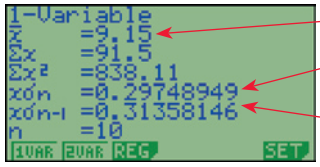


3 Press F2 **(CALC)**. You may need to first press F6 for more options.



4 Press F6 **(SET)**. Check that **1Var Xlist** is set to **List 1** and **1Var Freq** is set to **1**.

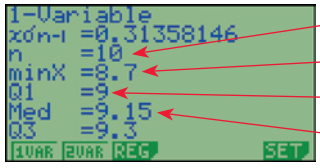
5 Press EXE to return to the previous screen, and then press F1 **(1Var)**. All statistics will now be on display using the scroll function.



Mean

Population standard deviation

Sample standard deviation

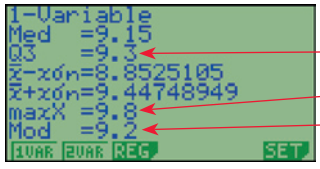


Number of scores

Lowest score

Lower quartile

Median



Upper quartile

Highest score

Mode

a The mean is denoted by the symbol $\bar{x}$.

a $\bar{x} = 9.15$

b The median is denoted by **Med**.

b Median = 9.15

c The mode is displayed by **Mod**. Check the scores for yourself as the data is bimodal only the largest mode is displayed.

c Mode = 9.0 and 9.2

d The range is the highest score (**maxX**) minus the lowest score (**minX**).

d Range = 9.8 – 8.7
= 1.1

e The interquartile range is the upper quartile (**Q3**) minus the lower quartile (**Q1**).

e Interquartile range = 9.3 – 9
= 0.3

f The population standard deviation is denoted by $x\sigma_n$.

f $\sigma_n \approx 0.297$

g The sample standard deviation is denoted by $x\sigma_{n-1}$.

g $\sigma_{n-1} \approx 0.314$

Having identified that the mean and median are measures of location and that range, interquartile range and standard deviation are measures of spread, it is important that you can recognise the effect that the members of a set have on these measures.

Consider the case of a basketball team. There are five players on the team, whose heights are: 1.91 m, 1.85 m, 1.52 m, 1.93 m and 1.99 m.

The team's mean height is 1.84 m. Only one of the five players in the team is shorter than the mean height. This is because there is one member of the data set whose height is much less than the others. A score in a data set that is either much less or much greater than all others is called an **outlier**.

An outlier will either reduce or increase the mean such that the mean is no longer typical of the data set. In such cases, the median is a better measure of location than the mean.

WORKED EXAMPLE 4

In a small street there are five houses. The values of these houses are: $450 000, $465 000, $465 000, $480 000, $495 000.
A new house is built and valued at $750 000. Describe the effect that this outlier has on the:
a mean b median c mode (the score that occurs most often).

THINK	WRITE
a 1 Calculate the mean before the new house is built.	a Before new house is built: Total = $2 355 000 Mean = $2 355 000 ÷ 5 = $471 000
2 Calculate the mean after the new house is built.	After new house is built: Total = $3 105 000 Mean = $3 105 000 ÷ 6 = $517 500
3 Comment on the change in the mean caused by the outlier.	The outlier has caused the mean to increase by $46 500. Only the new house is valued at more than the mean and, as such, has made the mean a poor measure of the typical price.
b 1 Calculate the median before the new house is built.	b Before new house is built: Median = $465 000
2 Calculate the median after the new house is built.	After new house is built: Median = ($465 000 + $480 000) ÷ 2 = $472 500

	3	Comment on the change in the median caused by the outlier.	The outlier has caused only a small increase in the median and, as such, the median remains a good measure of the typical score in this data set.
c	1	Calculate the mode before the new house is built.	c Before new house is built: Mode = $465 000
	2	Calculate the mode after the new house is built.	After new house is built: Mode = $465 000
	3	Comment on the change in the mode caused by the outlier.	The outlier has had no effect on the mode.

Generally the mean is the most vulnerable measure of location when an outlier is added to a data set. The median is affected only by the addition of the extra score and is not affected by the size of that score. The outlier will have no effect on the mode.

REMEMBER

1. The mean and median are measures of location in a data set.
   - The mean is calculated by adding the scores and then dividing by the number of scores. The mean is calculated using the formulas:
   $$\bar{x} = \frac{\Sigma x}{n} \quad \text{or} \quad \bar{x} = \frac{\Sigma fx}{\Sigma f}$$
   - The median is the middle score or the average of the two middle scores in a data set.
2. The range, interquartile range and standard deviation are measures of spread.
   - The range is the difference between the highest and lowest scores.
   - The interquartile range is the difference between the upper and lower quartiles.
   - The standard deviation is found using the σ_n (population) or s_n (sample) functions on the calculator.
3. An outlier is a score in a data set that is either much less or much greater than all other scores in the set.
4. All important summary statistics can be found by entering data into a graphics calculator.

EXERCISE
4A Measures of location and spread

eBookplus
Digital docs
SkillSHEET 4.1
doc-1330
Finding the mean
SkillSHEET 4.2
doc-1331
Finding the mode

1 WE1 The number of goals scored by a team in 10 games of soccer are:
2, 1, 3, 1, 0, 0, 1, 1, 6, 1.
- **a** Calculate the mean number of goals scored.
- **b** Calculate the median number of goals scored.

2 For each of the following sets of scores, calculate the mean, median and mode (if one exists).
- **a** 56, 75, 88, 20, 37, 23, 44
- **b** 2, 1, 7, 4, 6, 1, 1, 4, 5, 3
- **c** 9.9, 9.4, 9.8, 9.6, 9.0, 9.2, 9.8, 9.9
- **d** 13, 15, 16, 17, 10, 13, 15, 14, 19, 20

eBookplus

Digital doc
SkillSHEET 4.3
doc-1332
Finding the median

3 The table at right shows the scores out of 10 by a class of 30 students on a spelling test.

a Use the statistics function on your calculator to find the mean score.

b Add a cumulative frequency column to the table and use it to calculate the median score.

c State the mode.

Score	Frequency
4	2
5	6
6	7
7	9
8	3
9	2
10	1

eBookplus

Digital docs
EXCEL Spreadsheet
doc-1333
One variable statistics
GC program — Casio
doc-1334
UV Stats
GC program — TI
doc-1335
UV Stats

4 The table below shows the scores achieved by a football team over a season.

Score	Class centre	Frequency	Cumulative frequency
1–5		2	
6–10		4	
11–15		8	
16–20		7	
21–25		3	
26–30		1	

a Copy and complete the table.

b Calculate the mean.

c Draw a cumulative frequency histogram and curve and use them to estimate the median.

eBookplus

Digital doc
SkillSHEET 4.4
doc-1336
Finding the range

5 WE2 Below is the number of students in each class at a small primary school.
28, 29, 27, 28, 30, 28, 25, 27, 23, 28, 27, 28

a Calculate the range of the distribution.

b Calculate the interquartile range.

c Use the statistics function on your calculator to find the mean and standard deviation.

eBookplus

Digital docs
SkillSHEET 4.5
doc-1337
Finding the interquartile range
SkillSHEET 4.6
doc-1338
Choosing the appropriate standard deviation
EXCEL Spreadsheet
doc-1339
Boxplots

6 The table below shows the number of cars sold in a car yard each week over one year.

No. of cars	Frequency
6	3
7	5
8	9
9	15
10	11
11	8
12	1

a Calculate the range of the number of cars sold.

b Add a cumulative frequency column to the table and use the table to calculate:

i the median **ii** the upper and lower quartiles

iii the interquartile range.

c Use the statistics function on the calculator to find:

i the mean **ii** the standard deviation.

d Draw a box-and-whisker plot of the data.

7 The table below shows crowds at each match for a team during football season.

Crowd	Class centre	Frequency	Cumulative frequency
10 000–15 000		5	
15 000–20 000		8	
20 000–25 000		6	
25 000–30 000		4	
30 000–35 000		3	

a Copy and complete the table.
b Draw a cumulative frequency histogram and curve.
c Use the graph in part **b** to estimate the interquartile range.
d Find the mean and standard deviation. (Give your answer correct to 2 significant figures.)

8 WE3 Below are the scores of two rugby league teams over a period of 10 matches.
Team A: 14, 16, 16, 20, 10, 12, 18, 16, 18, 20
Team B: 28, 12, 32, 2, 0, 8, 40, 10, 12, 16
a For each team calculate the mean score.
b For each team calculate:
i the range **ii** the interquartile range **iii** the standard deviation.
c Comment on the difference between the performance of the two teams over this 10-game period.

The information below is to be used for questions 9 to 12.
A basketball squad has eight players. The mean height of the eight players is 1.8 m, and the standard deviation in the heights of the players is 0.1 m. In the first game the tallest player, who is 1.9 m tall, is injured and replaced in the squad by a player who is 1.98 m tall.

9 MC The mean height of the basketball squad will now be:
A 1.8 m **B** 1.81 m **C** 1.86 m **D** 1.96 m

10 MC As a result of the substitution:
A the standard deviation will increase
B the standard deviation will decrease
C the standard deviation will be unchanged
D the effect on the standard deviation cannot be calculated

11 MC As a result of the substitution:
A the range will increase
B the range will decrease
C the range will be unchanged
D the effect on the range cannot be calculated

12 MC As a result of the substitution:
A the interquartile range will increase
B the interquartile range will decrease
C the interquartile range will be unchanged
D the effect on the interquartile range cannot be calculated

13 WE4 James recorded the following five marks on his Maths tests during the year:
78, 77, 80, 85 and 80.
a Calculate:
i the mean **ii** the median **iii** the mode.
b In James' final exam he scored only 20. For the six test results calculate:
i the mean **ii** the median **iii** the mode.
c Describe the effect that the outlier had on the mean, median and mode.

14 The mean of a set of five scores is 60. A score of 90 is added to the data set. Describe the effect that this outlier will have on the mean.

15 MC Julie is currently in Year 12. The table below shows the number of days that Julie has been absent from school in each of the previous five years.

Year	No. of days absent
7	0
8	1
9	3
10	2
11	0

During Year 12, Julie became seriously ill and was forced to have 37 days off school. According to statistics calculated on Julie's absences over six years, this outlier will have the greatest effect on:

A the mean **B** the median **C** the mode **D** all of the above

16 A small company has four employees who each earn $397.50 per week. Later, a manager is employed who earns $1645.00 per week.

a Calculate the mean, median and mode wages.

b What effect does the manager's wage have on the:

i mean? **ii** median? **iii** mode?

c A wage debate is conducted with the employees asking for a rise. Would the mean, median or mode be quoted:

i in support of a wage rise by the employees?

ii against a wage rise by the employer?

Explain your answers.

Further development

17 a On a particular day, the number of cars that stopped at the drive-in area at a McFast restaurant during each hour (from 8.00 am until 11.00 pm) is shown below.
15 10 4 13 12 28 25 15 19 29 24 22 25 30 15
Find the interquartile range of this set of data.

b At the nearby Kirby's Fried Chicken restaurant on the same day, the number of cars stopping during each hour that the restaurant was open is shown below.
6 10 15 22 18 11 13 14 21 17 25 20 19 10 17
Find the interquartile range of these data.

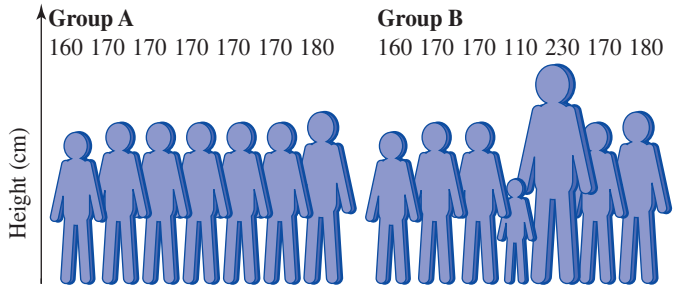
c What do these values suggest about the two restaurants?

18 Consider the following two groups of people.

a Calculate the mean height, median height and mode height for each group. What do you notice?

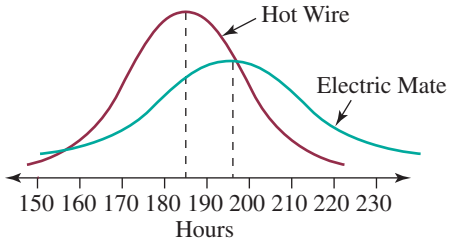
b Are the groups really the same?

c Which group would you expect to show the greatest range in heights?



d Which group would you expect to show the greatest interquartile range in heights?
e Which group would you expect to show the greatest standard deviation in heights?
f Calculate these statistics to confirm your predictions.

19 The curves drawn at right show the lifetimes of two samples of different brands of toaster elements when subject to continued use.



a Which brand has the longer mean life?
b Estimate the mean life of each brand.
c Which brand has the greater standard deviation in its performance?
d What does this say about the consistency of this element?
e Which brand is better? Give a case in support of Electric Mate. Give a case in support of Hot Wire.

20

Position	Salary ($)	Number of employees
Machine operator	19 000	50
Machine mechanic	21 000	15
Floor steward	25 000	10
Manager	63 000	4
Chief executive officer	81 000	1

a Workers are arguing for a pay rise but the management of the factory claims that workers are well paid because the mean salary of the factory is $23 100. Are they being honest?
b Suppose that you were representing the factory workers and had to write a short submission in support of the pay rise. How could you explain the management's claim? Quote some other statistics in favour of your case.

21 At a particular point of the highway the speed limit is 60 km/h. A speed camera records the following speeds:

58, 57, 65, 62, 57, 55, 40, 44, 56, 59

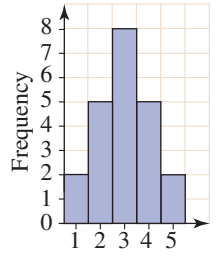
At a second point on the highway where the speed limit is 100 km/h, a speed camera takes the following measurements:

89, 91, 87, 130, 75, 88, 115, 145, 84, 90

With reference to both measures of location and spread discuss at which location speeding is the greater problem.

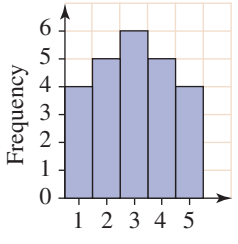
4B Skewness

By looking at a graph, we can make judgements about the nature of a data set. Consider the first graph shown on the right.



This graph is symmetrical and we can see that the mean, median and mode are all equal to 3. The majority of scores are clustered around the mean. This is an example of a normal distribution.

We can compare the standard deviation of data sets by looking at such graphs. The more clustered the data set, the smaller the standard deviation.

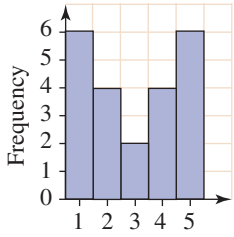


The second graph is still normally distributed with the mean, median and mode still equal to 3. However, there are more scores which are further away from the mean and, hence, the standard deviation of the data set is greater.

The third graph shows a data set where the scores are not clustered and there are two modes at either end of the distribution.

In this example, although it is still symmetrical there are two modes, 1 and 5, while the mean and median are still 3. The standard deviation in this distribution is greater than either of the two previous examples as there are more scores further away from the mean.

The mean and median can be seen from the graph only because it is symmetrical.



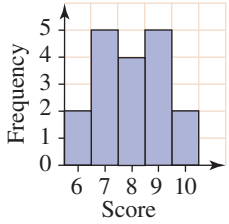
WORKED EXAMPLE 5

The figure on the right shows the distribution of a set of scores on a spelling test.

a Is the graph symmetrical?

b What is the mode(s)?

c Can the mean and median be seen from the graph?



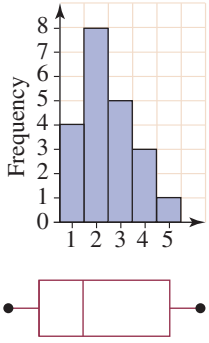
eBook*plus*

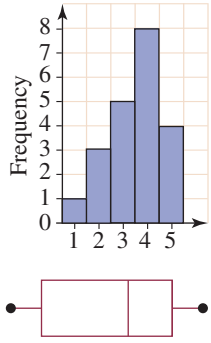
Tutorial
int-2418
Worked example 5

THINK	WRITE
a The columns either side of the middle are equal.	a The graph is symmetrical.
b The scores that occur the most often are 7 and 9.	b Mode = 7 and 9
c The middle score will be the mean and median.	c Mean = 8, median = 8

When a graph is not symmetrical, the mean and median cannot be easily seen from the graph. Consider the distribution in the graph below left.

The way in which the data are gathered to one end of the distribution is called the **skewness**. A greater number of scores are distributed at the lower end of the distribution. In this case, the data are said to be positively skewed. Similarly, when most of the scores are distributed at the upper end, the data are said to be negatively skewed, as shown in the graph below right.



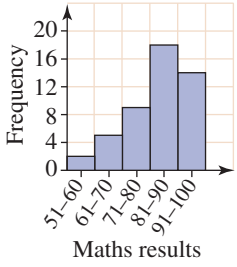


WORKED EXAMPLE 6

The distribution on the right shows the results of the Maths Trial HSC at a certain school.

a What is the modal class?

b Describe the skewness of the data set shown on the right.



THINK	WRITE
a The class occurring the most often is the 81–90 class.	**a** Modal class = 81 – 90
b The majority of data are at the upper end of the distribution.	**b** The data are negatively skewed.

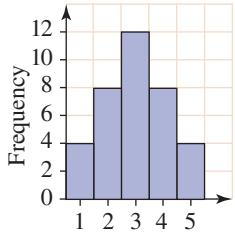
REMEMBER

1. A distribution is symmetrical when the data are equally distributed around the mean.
2. When the data are symmetrical, the median and mean will both be the middle score.
3. When the data are clustered around the mean, the standard deviation is smaller.
4. When the majority of scores are at the lower end of a distribution, it is said to be positively skewed.
5. When the majority of scores are at the upper end of the distribution, it is said to be negatively skewed.

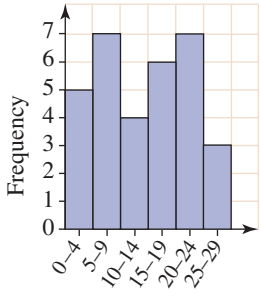
EXERCISE 4B

Skewness

1 WE5 In the distribution on the right:
- **a** is the graph symmetrical?
- **b** what is the modal class(es)?
- **c** can the mean and median be seen from the graph? and, if so, what are their values?



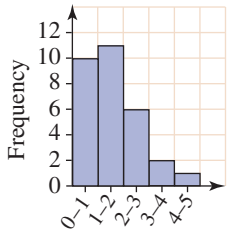
2 For the distribution shown on the right:
- **a** are the data symmetrical?
- **b** what is the modal class(es)?
- **c** can the mean and median be seen from the graph? and, if so, what are their values?



3 The table on the right shows the number of goals scored by a hockey team throughout a season.
- **a** Show this information in a frequency histogram.
- **b** Are the data symmetrical?
- **c** What is the mode(s)?
- **d** Can the mean and median be seen for this distribution? and, if so, what are their values?

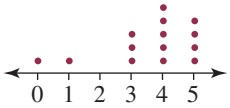
No. of goals	Frequency
0	6
1	4
2	4
3	4
4	4
5	6

4 WE6 For the distribution shown on the right:
- **a** what is the modal class(es)?
- **b** describe the skewness of the distribution.

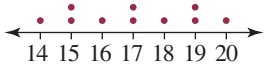


5 For each of the following dot plots describe the skewness of the distribution.

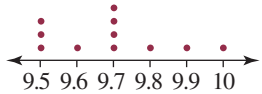
a



b



c



6 For the stem-and-leaf plots drawn below describe the distribution

a Key 3|5 = 35

Stem	*Leaf*
2	2 5 9
3	0 0 1 2 5 8 9
4	2 2 8 9
5	0 9
6	0

b Key 4|3 = 4.3 4*|6 = 4.6

Stem	*Leaf*
2*	9
3	0 4
3*	5 5 8 8
4	0 0 0 1 1 3 4 4
4*	5 5 5 6 7 8 8 9 9

7 The table below shows the number of goals scored by a basketball team throughout a season.

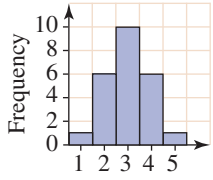
No. of goals	Frequency
11–20	3
21–30	6
31–40	7
41–50	23
51–60	21

a Draw a frequency histogram of the data.

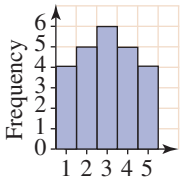
b Describe the data set in terms of its skewness.

8 MC Which of the distributions below has the smallest standard deviation?

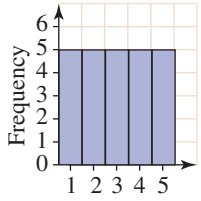
A



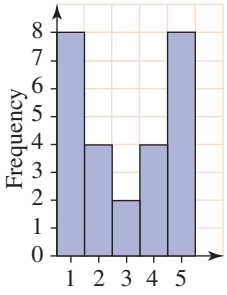
B



C



D



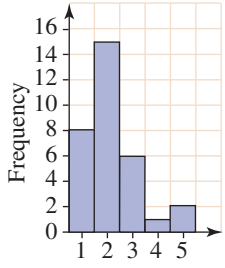
9 MC The distribution represented by the graph on the right is:

A positively skewed

B negatively skewed

C symmetrical

D normally distributed



10 A movie is shown at a cinema 30 times during the week. The number of people attending each session of the movie is shown in the table below.

No. of people	Frequency
1–50	2
51–100	3
101–150	5
151–200	10
201–250	10

a Present the data in a frequency histogram.
b Are the data symmetrical?
c What is the modal class(es)?
d Describe the skewness of the distribution.

11 Year 12 at Wallarwella High School sit exams in Chemistry and Maths. The results are shown in the table below.

Mark	Chemistry	Maths
31–40	2	3
41–50	9	4
51–60	7	6
61–70	4	7
71–80	7	9
81–90	9	7
91–100	2	4

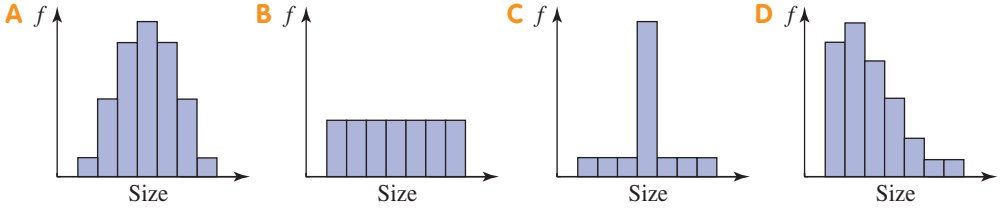
a Is either distribution symmetrical?
b If either distribution is not symmetrical, state whether it is positively or negatively skewed.
c State the mode of each distribution.
d In which subject is the standard deviation greater? Explain your answer.

eBookplus
Digital doc
WorkSHEET 4.1
doc-1340

12 Draw an example of a graph which is:
a symmetrical
b positively skewed with one mode
c negatively skewed with two modes.

Further development

13 MC Match the boxplot with its most likely histogram.



14 For each of the following, write down whether the mean or the median would provide a better indication of the centre of the distribution.

a A positively skewed distribution
b A symmetric distribution
c A distribution with an outlier
d A negatively skewed distribution

15 For each of the following, state whether you would expect the distribution to be positively skewed, negatively skewed or symmetrical.

a The number of days absent had by students of a school in one term
b The height of Year 12 students
c The number of cars in each household
d The scores out of 10 by Year 12 students on a times tables quiz

4C Displaying multiple data sets

Two data sets can be compared using a number of the displays that have been studied in earlier parts of this course. Presenting both sets of data on the same display gives a quick and easy comparison.

Stem-and-leaf plots

Two sets of data can be displayed on the same stem-and-leaf plot. This is done by having the stem in the centre of the plot, with both sets of data back to back.

WORKED EXAMPLE 7

The data shown below display the marks of 15 students in both English and Maths.

English:	**45**	**67**	**81**	**59**	**66**	**61**	**78**	**71**	**74**	**91**	**60**	**49**	**58**	**62**	**70**
Maths:	**85**	**71**	**49**	**66**	**64**	**68**	**75**	**71**	**69**	**60**	**63**	**80**	**87**	**54**	**59**

Display the data in a back-to-back stem-and-leaf plot.

THINK

1. Write a key at the top of the stem-and-leaf plot.
2. Draw the stem showing categories of 10 in the centre of the page.
3. Display the information for English on the left of the stem.
4. Display the information for Maths on the right of the stem.

WRITE

Key: 4|5 = 45

English		Maths
9 5	**4**	9
9 8	**5**	4 9
7 6 2 1 0	**6**	0 3 4 6 8 9
8 4 1 0	**7**	1 1 5
1	**8**	0 5 7
1	**9**	

This stem-and-leaf plot allows for both distributions to be easily seen, and for a judgement on the skewness of the distribution to be made.

Box-and-whisker plots

A single scale can be used. Draw two box-and-whisker plots on that scale that will allow the comparison of the median, range and interquartile range of two distributions to be compared.

WORKED EXAMPLE 8

Use the back-to-back stem-and-leaf plot drawn in worked example 7 to:

a calculate the median of each distribution
b calculate the range of each distribution
c calculate the interquartile range of each distribution
d draw a box-and-whisker plot of each distribution on the same scale.

THINK

Method 1: Technology-free

a The median will be the eighth score in each distribution.

b To calculate the range of each distribution, subtract the lowest score from the highest score.

c 1 The lower quartile will be the fourth score.

2 The upper quartile will be the twelfth score.

3 The interquartile range is the difference between the quartiles.

d 1 Draw a scale.

2 Draw the English box-and-whisker plot.

3 Draw the Maths box-and-whisker plot.

Method 2: Technology-enabled

1 From the **MENU** select **STAT**.

2 Enter the data for English in **List 1** and the data for Maths in **List 2**.

3 Press F1 **(GRPH)**, and then F6 **(SET)**. We will set the English data as **GPH1** and the Maths data as **GPH2**. Press F1 **(GPH1)** and enter the settings shown in the screen at right.

4 Press EXE once these settings have been entered.
Press F6 **(SET)** and F2 **(GPH2)**, and again enter the settings shown at right.

5 To show both graphs on the same screen press EXE after entering the settings above, press F4 **(SEL)**, and set both **StatGraph1** and **StatGraph2** to **DrawOn** as shown.

WRITE

a English median = 66
Maths median = 68

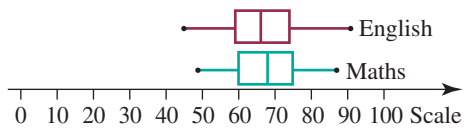
b English range = 91 – 45
= 46
Maths range = 87 – 49
= 38

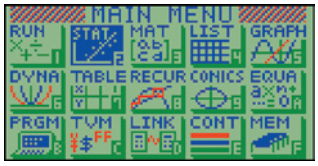
c English lower quartile = 59
Maths lower quartile = 60

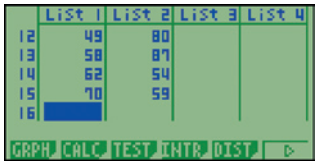
English upper quartile = 74
Maths upper quartile = 75

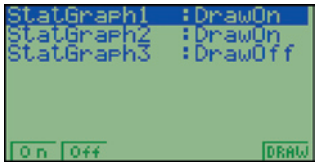
English interquartile range = 74 – 59
= 15
Maths interquartile range = 75 – 60
= 15

d

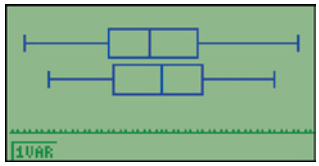




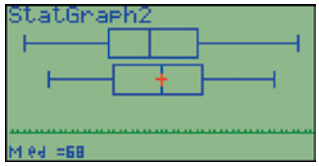




6 Press F6 **(DRAW)** to draw both graphs.



7 Press **SHIFT** F1 **(TRACE)**, and use the arrow keys to move around to the five key points on each graph. The screen at right displays the median for the Maths data.



Examining exam results

Collect data on the most recent exam that has been done in your class.

1 Display the data in a stem-and-leaf plot.
2 Find all the information needed to display the data in a box-and-whisker plot.
3 Is there any skewness evident in the data?
4 Which measure of location best describes the typical score in this data set?

WORKED EXAMPLE 9

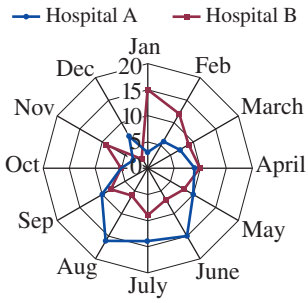
The table below shows the number of admissions to two hospitals, each month, over a one-year period. Display both sets of data on a radar chart.

Month	Hospital A	Hospital B
January	3	15
February	6	12
March	7	9
April	9	10
May	10	8
June	15	7
July	14	9
August	16	6
September	10	8
October	5	5
November	3	9
December	7	2

THINK

1 Draw the radar with a 30° angle between the months.
2 Draw a scale around the radar.
3 Plot each set of points.

WRITE



Area charts are another method of comparing information. In an area chart, line graphs are stacked on top of each other, thus allowing the area between each line graph to serve as the comparison between the data sets.

WORKED EXAMPLE 10

eBookplus
Tutorial
int-2419
Worked example 10

The table below shows the amount of rainfall, in millimetres, in Sydney, Melbourne and Brisbane each month throughout a year.

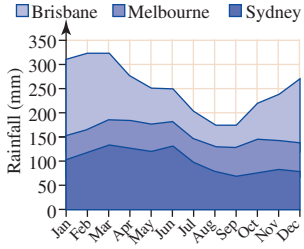
	January	February	March	April	May	June
Sydney	103	117.1	133.7	126.6	120.4	131.7
Melbourne	49	47.7	51.8	58.4	57.2	50.2
Brisbane	159.6	158.3	140.7	92.5	73.7	67.8
	July	**August**	**September**	**October**	**November**	**December**
Sydney	98.2	79.8	69.9	77.5	83.1	79.6
Melbourne	48.7	50.6	59.4	67.7	60.2	59.9
Brisbane	56.5	45.9	45.7	75.4	97	133.3

Show this information in an area chart.

THINK

1. Draw a pair of axes. The vertical axis will need to be at least the rainfall total of all three cities in the wettest month.
2. Draw a line graph of Sydney's rainfall, and shade the area below it.
3. Add Melbourne's rainfall to Sydney's rainfall, and draw a line graph showing these figures. Colour the area between the two graphs, as this area represents Melbourne's rainfall.
4. Add Brisbane's rainfall to the previous total. Colour the area above the previous line, as this area represents Brisbane's rainfall.

WRITE



REMEMBER

1. Data can be compared by showing two sets of data on the same display.
2. Two sets of data are shown on a stem-and-leaf plot by displaying the data back to back.
3. Two box-and-whisker plots can be drawn on the same scale to compare the ranges, interquartile ranges and medians.
4. A radar chart can be used to compare trends over a period of time by plotting two sets of data on one radar chart.
5. An area graph can be used to compare multiple sets of data. The area in each section of the graph allows for comparison between quantities.

EXERCISE

4C Displaying multiple data sets

eBookplus
Digital doc
SkillSHEET 4.7
doc-1341
Compiling a stem-and-leaf plot

1 WE7 In a class of 30 students there are 15 boys and 15 girls. Their heights are measured and are listed below.
Boys: 1.65, 1.71, 1.59, 1.74, 1.66, 1.69, 1.72, 1.66, 1.65, 1.64, 1.68, 1.74, 1.57, 1.59, 1.60
Girls: 1.66, 1.69, 1.58, 1.55, 1.51, 1.56, 1.64, 1.69, 1.70, 1.57, 1.52, 1.58, 1.64, 1.68, 1.67
Display this information in a back-to-back stem-and-leaf plot.

2 The number of points scored in each match by two rugby union teams are shown below.
Team 1: 34, 32, 24, 25, 8, 18, 17, 23, 29, 40, 19, 42
Team 2: 23, 20, 35, 21, 46, 7, 9, 24, 27, 38, 41, 30
Display these data in a back-to-back stem-and-leaf plot.

eBookplus
Digital doc
SkillSHEET 4.8
doc-1342
Finding the mean, median, mode from a stem-and-leaf plot

3 WE8 The stem-and-leaf plot below is used to display the number of vehicles sold by the Ford and Holden dealerships in a Sydney suburb each week for a three-month period.
Key: 1|5 = 15

Ford		Holden
7 4	**0**	3 9
9 5 2 2 1 0	**1**	1 1 1 6 6 8
8 5 4 4	**2**	2 2 7 9
0	**3**	5

a State the median of both distributions.
b Calculate the range of both distributions.
c Calculate the interquartile range of both distributions.
d Show both distributions on a box-and-whisker plot.

eBookplus
Digital doc
Skillsheet 4.9
doc-1343
Drawing a box-and-whisker plot

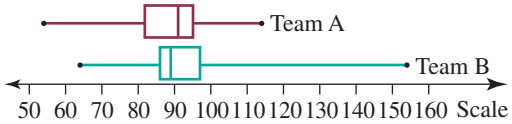
4 A motoring organisation tests two different brands of tyres. Twenty tyres of each brand are tested to find out the number of kilometres each tyre could travel before the tread had worn down. The results are shown in the stem-and-leaf plot below.
Key: 1|2 = 12 000 km 1*|7 = 17 000 km

Brand A		Brand B
9 8	**0***	
4 3 1 1 0	**1**	0 0 1 1 2 2 4
7 7 7 6 6 5	**1***	5 6 7 8 8 8 9
4 4 3 1 1 0 0	**2**	0 1 3 4
	2*	5 5

Draw two box-and-whisker plots on the same scale to display this information.

5 The figures below show the ratings of two radio stations each week over a three-month period.
Station A: 9.2, 9.4, 9.2, 9.5, 9.7, 9.9, 10.1, 9.1, 8.8, 8.7, 9.0, 8.5, 9.3
Station B: 8.5, 8.1, 8.2, 8.9, 9.0, 9.2, 8.4, 8.7, 8.8, 10.5, 11.2, 11.4, 8.7
a Display the information in a back-to-back stem-and-leaf plot.
b Use the stem-and-leaf plot to display both sets of data on the same box-and-whisker plot.

6 The box-and-whisker plot drawn on the below displays statistical data for two AFL teams over a season.

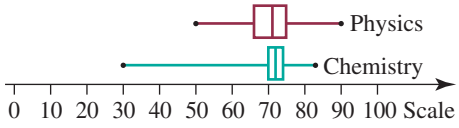


a Which team had the higher median score?
b What was the range of scores for each team?
c For each team calculate the interquartile range.

7 The two five-number summaries below show the performance of Emad and Larry on their Mathematics exams throughout the year.
Emad: 45, 64, 68, 76, 80
Larry: 51, 58, 65, 72, 75

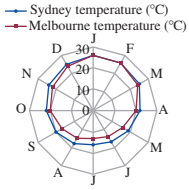
a Compare the performance of Emad and Larry on a box-and-whisker plot.
b What is the range for both students?
c What is the interquartile range for both students?

8 MC The box-and-whisker plot drawn on the below shows Emma's performance in her Physics and Chemistry exams. Which of the following statements is correct?



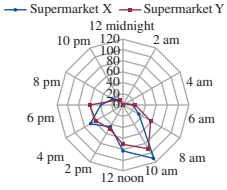
A The median of Emma's mark in Physics is greater than for Chemistry.
B The range of Emma's marks in Physics is greater than in Chemistry.
C The interquartile range of Emma's marks in Physics is greater than in Chemistry.
D All of the above.

9 WE9 This radar chart shows the average daily maximum temperature for both Sydney and Melbourne for each month of a year.



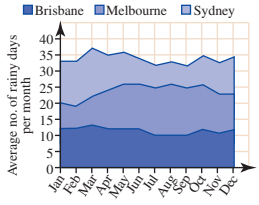
a Which month had the lowest temperature in Sydney?
b What was the range of temperatures in Melbourne?
c What was the average of the temperatures in Sydney?

10 This radar chart shows the number of customers in two different supermarkets at two-hour intervals.



a Find the range for each supermarket.
b Describe the general pattern at each supermarket.

11 WE10 This area chart shows the average number of rainy days each month in Sydney, Melbourne and Brisbane. Display this information as a table.



Further development

12 The stem-and-leaf plot drawn below shows the marks obtained by 20 students in both English and Maths.
Key: 7 | 1 = 71

English		Maths
	4	1 7
7 4 1 0	**5**	2 4 7 9 9
9 9 7 6 6 5 3 1 1 0	**6**	1 3 3 4 6 6
8 7 7 5 2	**7**	4 4 4 8
2	**8**	3 6
	9	4

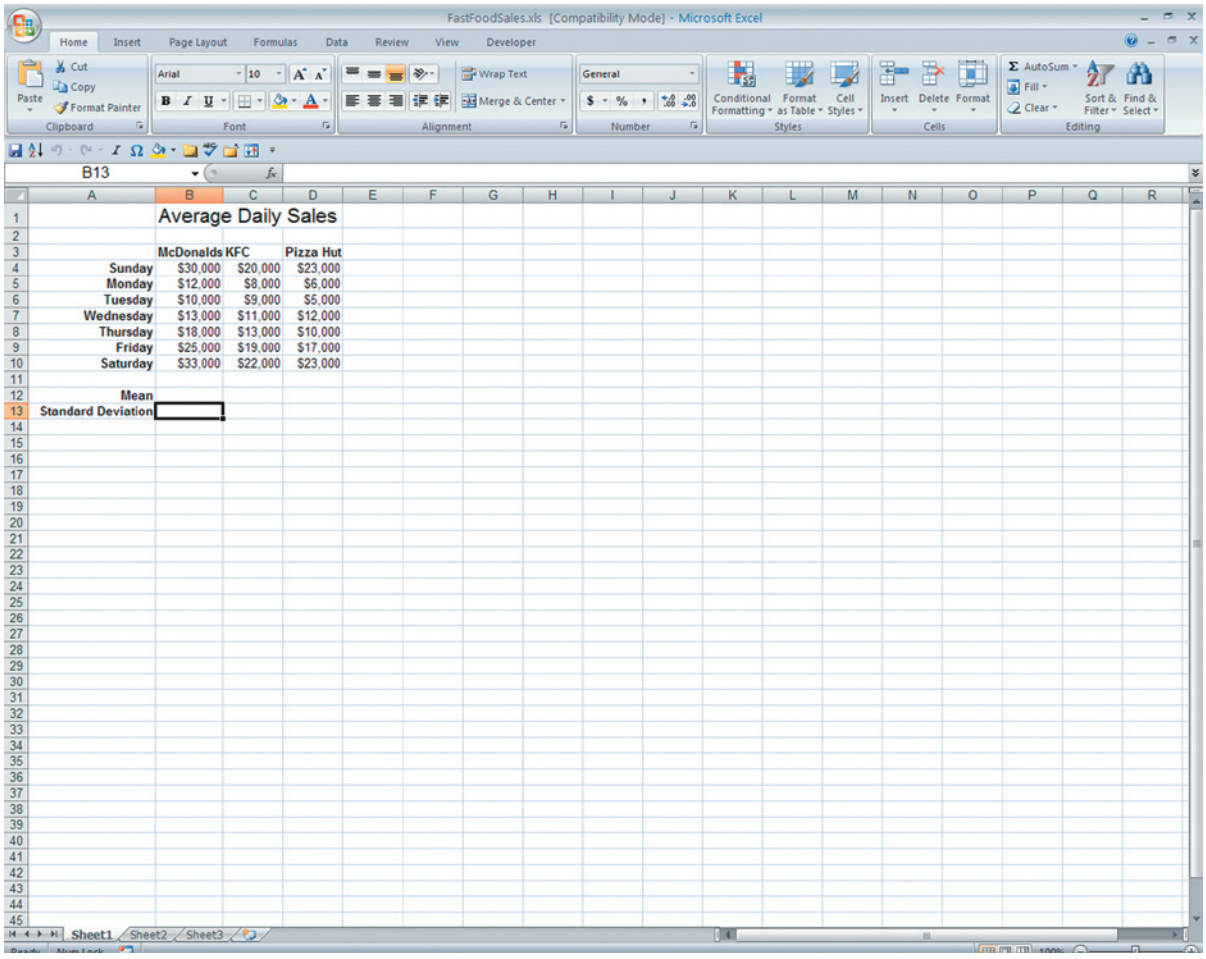
a Calculate the median mark for both English and Maths.
b Calculate the range of marks for both English and Maths.
c Comment on the distribution of marks in each of the subjects.

eBookplus

Digital doc
EXCEL Spreadsheet
doc-1344
Fast Food Sales

Computer Application 1: Displaying statistical data

1. From the Maths Quest General Mathematics HSC Course eBookPLUS access the spreadsheet *Fast Food Sales*.



2. In cell **B12** use the spreadsheet's inbuilt statistical function to find McDonald's average daily sales. [=**AVERAGE(B4:B10)**]
3. In cell **B13** use the spreadsheet's inbuilt statistical function to find the standard deviation of McDonald's daily sales. [=**STDEV(B4:B10)**]
4. Under **Edit**, use the **Fill** and **Right** functions to copy these formulas for KFC and Pizza Hut.
5. Use the charting facility to draw an area chart of the figures presented.

4D Comparison of data sets

When multiple data displays are used to display similar sets of data, comparisons and conclusions can then be drawn about the data.

Multiple displays such as stem-and-leaf plots and box-and-whisker plots allow for comparison of statistics such as the median, range and interquartile range, while radar charts and area charts allow for trends and overall quantities to be compared.

WORKED EXAMPLE 11

A bank surveys the average morning and afternoon waiting time for customers. The figures were taken each Monday to Friday in the morning and afternoon for one month. The stem-and-leaf plot at right shows the results.

Key: 1|2 = 1.2 minutes

Morning		Afternoon
7	**0**	7 8 8
8 6 3 1 1	**1**	1 1 2 4 4 5 6 6 6 7
9 6 6 6 5 5 4 3 3 1	**2**	2 5 5 8
9 5 2	**3**	1 6
5	**4**	
	5	7

a Find the median morning waiting time and the median afternoon waiting time.

b Calculate the range for morning waiting times and the range for afternoon waiting times.

c What conclusions can be made from the display about the average waiting time at the bank in the morning compared with the afternoon?

THINK

a There are 20 scores in each set and so the median will be the average of the 10th and 11th scores.

b For each data set, subtract the lowest score from the highest score.

c Conclude that waiting time in the afternoon is generally less and more consistent except for one outlier.

WRITE

a Morning: Median = (2.4 + 2.5) ÷ 2
= 2.45 minutes
Afternoon: Median = (1.6 + 1.6) ÷ 2
= 1.6 minutes

b Morning: Range = 4.5 – 0.7
= 3.8 minutes
Afternoon: Range = 5.7 – 0.7
= 5 minutes

c The waiting time is generally shorter in the afternoon. There is one outlier in the afternoon scores which causes the range to be larger. However, apart from this outlier the afternoon scores are less spread.

Two-way tables can also be a meaningful way of displaying data. A two-way table allows for two variables to be compared.

WORKED EXAMPLE 12

A survey of 25 000 people is taken. The sex of each respondant is noted and whether they are a smoker or non-smoker is also noted. The results are displayed in the two-way table below.

	Males	Females	Totals
Smokers	4 125	4 592	8 717
Non-smokers	8 436	7 847	16 283
Totals	12 561	12 439	**25 000**

a What percentage of the females surveyed were smokers?

b What percentage of the smokers surveyed were female?

THINK	WRITE
a Write 4592 as a percentage of 12 439.	**a** Percentage of females who smoke $= \frac{4592}{12\,439} \times 100\%$ $= 36.9\%$
b Write 4592 as a percentage of 8717.	**b** Percentage of smokers who are female $= \frac{4592}{8717} \times 100\%$ $= 52.7\%$

The most common method, however, for comparing data sets is to compare the summary statistics from the data sets. The measures of location such as mean and median are used to compare the typical score in a data set. Measures of spread such as range, interquartile range and standard deviation are used to make assessments about the consistency of scores in the data set.

WORKED EXAMPLE 13

eBook*plus*
Tutorial
int-2420
Worked example 13

Below are the scores for two students in eight Mathematics tests throughout the year.
Jane: 45, 62, 64, 55, 58, 51, 59, 62
Pierre: 84, 37, 45, 80, 74, 44, 46, 50

a Use the statistics function on the calculator to find the mean and standard deviation for each student.
b Which student had the better overall performance on the eight tests?
c Which student was more consistent over the eight tests?

THINK	WRITE
a Enter the statistics into your calculator and use the $\bar{x}$ function for the mean and the σ_n function for the standard deviation.	**a** Jane: $\bar{x} = 57$, $\sigma_n = 6$ Pierre: $\bar{x} = 57.5$, $\sigma_n = 17.4$
b The student with the higher mean performed better overall.	**b** Pierre performed slightly better overall, as his mean mark was higher than Jane's.
c The student with the lower standard deviation was more consistent.	**c** Jane was the more consistent student, as her standard deviation was much lower than Pierre's.

REMEMBER

1. When multiple displays are used for two or more sets of data, we can compare and contrast the data sets and determine if any relationship exists between them.
2. A multiple stem-and-leaf plot allows for a quick comparison of medians, ranges and interquartile ranges.
3. The summary statistics from two data sets can be compared quickly on a box-and-whisker plot.
4. Two-way tables can be used to make a comparison of data where two variables are involved.
5. The most commonly used comparisons are summary statistics to compare what is a typical score and what the spread of the data is.

EXERCISE

4D Comparison of data sets

1 WE11 The stem-and-leaf plot drawn below shows the marks obtained by 20 students in both English and Maths.

Key: 7|1 = 71

English		Maths
	4	1 7
7 4 1 0	**5**	2 4 7 9 9
9 9 7 6 6 5 3 1 1 0	**6**	1 3 3 4 6 6
8 7 7 5 2	**7**	4 4 4 8
2	**8**	3 6
	9	4

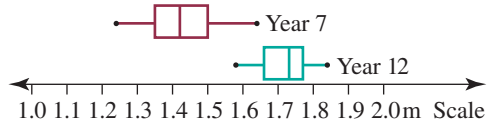
a Calculate the median mark for both English and Maths.
b Calculate the range of marks for both English and Maths.
c Comment on the distribution of marks in each of the subjects.

2 Tracey measures the heights of twenty Year 10 boys and twenty Year 10 girls and produces the following five-number summaries for each data set.

Boys: 1.47, 1.58, 1.64, 1.72, 1.81 Girls: 1.55, 1.59, 1.62, 1.66, 1.73

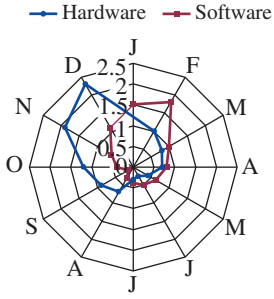
a Draw a box-and-whisker plot for both sets of data and display them on the same scale.
b What is the median of each distribution?
c What is the range of each distribution?
d What is the interquartile range for each distribution?
e Comment on the spread of the heights among the boys and the girls.

3 The box-and-whisker plots on the below show the heights of a sample of Year 7 boys and a similar-sized sample of Year 12 boys.



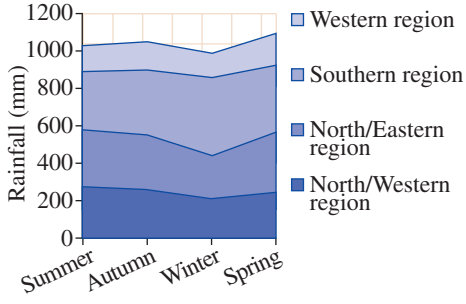
a Calculate the range of heights among both the Year 7 and Year 12 boys.
b Calculate the interquartile range of the heights among both the Year 7 and Year 12 boys.
c Comment on the relationship between the two data sets, both in terms of measures of location and measures of spread.

4 The values of hardware and software sales for a chain of computer stores are shown for each month in the radar chart on the right. Comment on any relationship observed in this chart between the sales of hardware and the sales of software.



5 The area chart on the right shows the rainfall in four areas of New South Wales throughout the year.

a Which region has the greatest rainfall?
b In which region is the range of rainfall figures least?
c What relationship exists between the rainfall in each of the areas?



6 WE12 The two-way table below shows the results of random breath testing by Sydney police over one weekend. A driver is charged if they record a reading of 0.05% prescribed concentration of alcohol (PCA).

	Males	Females	Totals
Over 0.05 PCA	26	7	33
Below 0.05 PCA	962	743	1705
Totals	988	750	**1738**

a What percentage of the drivers tested were female?
b What percentage of the drivers tested had a PCA over 0.05?
c What percentage of female drivers had a PCA over 0.05?
d What percentage of male drivers had a PCA over 0.05?
e Based on the above results, can any conclusion be drawn concerning the prevalence of drink driving among males and females? Explain your answer.

7 Ashley is the star player of a football team. To analyse the importance of Ashley to the team, the coach prepares the two-way table below showing the results of games over three years both when Ashley is playing and not playing.

	Won	Lost	Totals
Ashley playing	38	4	42
Ashley not playing	10	8	18
Totals	48	12	**60**

a What percentage of games were won when Ashley played?
b What percentage of games were won when Ashley did not play?
c Do you think that Ashley has a significant impact on the performance of the team? Explain your answer.

8 To compare the performance of city and country students in the HSC, the number of students achieving a UAI of at least 90 is studied in six city and six country high schools.

	City	Country	Totals
UAI ≥ 90	58	61	119
UAI < 90	551	569	1120
Totals	609	630	**1239**

a What percentage of city students achieved a UAI of at least 90?
b What percentage of country students achieved a UAI of at least 90?
c Of those students who achieved a UAI of at least 90, what percentage were from:
i the city? **ii** the country?
d Based on the above results, could any conclusion be drawn about the performance of city and country students in the HSC?

9 WE13 Calvin recorded his marks for each test that he did in Physics and Chemistry throughout the year.
Physics: 65, 74, 69, 66, 72, 64, 75, 60
Chemistry: 45, 85, 91, 42, 47, 72, 87, 85
a In which subject did Calvin achieve the better average mark?
b In which subject was Calvin more consistent? Explain your answer.

10 The police set up two radar speed checks in a country town. In both places the speed limit is 60 km/h. The results of the first 10 cars that have their speed checked are given below.
Point A: 60, 62, 58, 55, 59, 56, 65, 70, 61, 64
Point B: 55, 58, 59, 50, 40, 90, 54, 62, 60, 60

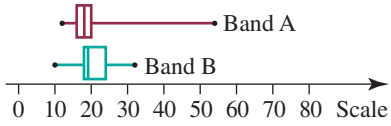
a Calculate the mean and standard deviation of the readings taken at each point.
b At which point are drivers generally driving faster?
c At which point is the spread of the readings taken greater? Explain your answer.

11 Aaron and Sunil open the batting for the local cricket team. The number of runs they have scored in each innings this season are listed below.

Aaron: 45, 43, 33, 56, 21, 38, 0, 29, 76, 40
Sunil: 5, 70, 12, 54, 68, 11, 8, 64, 32, 69

a Calculate the mean number of runs scored for each player.
b What is the range of runs scored by each player?
c What is the interquartile range of runs scored by each player?
d Which player would you consider to be the more consistent player? Explain your answer.

12 MC Andrea surveys the age of people attending a concert given by two bands. The box-and-whisker plot shown at right shows the results.



Which of the following conclusions could be drawn based on the above information?
A Band A attracts an older audience than Band B.
B Band A appeals to a wider age group than Band B.
C Band B attracts an older audience than Band A.
D None of the above.

13 MC Two drugs are tested to see which is more effective at fighting disease. The results are displayed in the two-way table below.

	Drug 1	Drug 2	Totals
Recovered	124	136	260
Not recovered	32	45	77
Totals	156	181	**337**

Of those patients who recovered, the percentage who were treated with drug 1 was:
A 46.3% B 47.7% C 69.0% D 79.5%

14 MC The figures below show the ages of the men's and women's champions at a tennis tournament.

Men's: 23, 24, 25, 26, 25, 25, 22, 23, 30, 24
Women's: 19, 27, 20, 26, 30, 18, 28, 25, 28, 22

Which of the following statements is correct?
A The mean age of the men's champions is greater than the mean age of the women's champions.
B The range is greater among the men's champions than among the women's champions.
C The interquartile range is greater among the men's champions than among the women's champions.
D The standard deviation is greater among the men's champions than among the women's champions.

eBook*plus*

Digital doc
WorkSHEET 4.2
doc-1345

15 A company producing matches advertise that there are 50 matches in each box. Two machines are used to distribute the matches into the boxes. The results from a sample taken from each machine are shown in the stem-and-leaf plot below.

Key: 5|1 = 51 5*|6 = 56

Machine A		Machine B
4	**4**	
9 9 8 7 7 6 6 5	**4***	5 7 8 9 9 9 9 9 9 9
4 3 2 2 2 2 1 1 1 0 0 0 0 0	**5**	0 0 0 0 0 1 1 1 1 1 2 2 3
5 5	**5***	9

a Display the data from both machines on a box-and-whisker plot.
b Calculate the mean and standard deviation of the number of matches distributed from both machines.
c Which machine is the more dependable? Explain your answer.

Further development

16 MC The two-way table has had one figure copied incorrectly.

	Men	Women	Total
Agree	45	92	137
Disagree	72	93	175
Total	117	195	

Which figure was copied incorrectly?
A The number of men who agreed.
B The number of men who disagreed.
C The number of women who agreed.
D The number of women who disagreed.

17 The two-way table shows the results of a study into whether men or women prefer to live alone.
Complete all missing information from the two-way table.

	Men	Women	Total
Live alone	12		35
Share with friends			
Total	21	39	

Developing a two-way table

Conduct a survey to determine the number of Year 12 students at your school who have their driver's licence. Use the results to complete the two-way table below.

	Males	Females	Totals
Licensed			
Unlicensed			
Totals			

a What percentage of females have their driver's licence?
b What percentage of licensed drivers are female?
c Are these figures the same?
d Design a data set where the percentage of females with their licence is equal to the percentage of licensed drivers who are female.
e What conditions are necessary for these two percentages to be equal?

SUMMARY

Measures of location and spread

- Measures of location give the typical score in the data set. The mean, median and mode are measures of location.
- The mean of a small data set is found using:

$$\bar{x} = \frac{\Sigma x}{n}$$

  where $\bar{x}$ = the mean, x = individual scores and n = number of scores.
- Where data is in a table, the mean is found using:

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$

  where $\bar{x}$ = the mean, x = individual scores and f = frequency.
- Measures of spread describe how spread out the data are. The range, interquartile range and standard deviation are measures of spread.
- An outlier is a single score that is much greater or much less than most of the scores. The outlier may have a great effect on the mean but has only a slight effect on the median and no effect on the mode in a small data set. The larger the data set, the less the effect a single outlier will have.

Skewness

- When the data are symmetrical, they are said to be normally distributed.
- The more clustered the data are around the mean, the smaller the standard deviation.
- When most of the data are below the mean, the data are said to be positively skewed.
- When most of the data are above the mean, the data are said to be negatively skewed.

Displaying multiple data sets

- Two sets of data can be displayed on a stem-and-leaf plot by displaying the data back to back.
- The summary statistics from two data sets can be displayed by using the same scale and drawing two box-and-whisker plots.
- Two sets of data can be displayed on a radar chart to display related trends over a period of time.
- An area chart can be drawn to display several sets of data. The area in each section of the graph then displays the quantities for comparison.

Comparison of data sets

- The summary statistics from two data sets can be compared from a stem-and-leaf plot or box-and-whisker plot.
- Two-way tables are used to compare data where there are two variables involved.
- Data are most commonly compared using the mean and standard deviation.

CHAPTER REVIEW

MULTIPLE CHOICE

1 The table below shows the number of patients seen each day by a local doctor.

No. of patients	Frequency
12	3
13	8
14	15
15	23
16	18
17	13

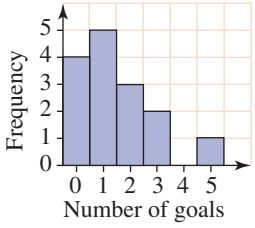
Which of the following statements are correct?

A The range of the data is 20.
B The mean of the data is 15.05.
C The standard deviation of the data is 1.34.
D The median of the data is 15.

2 The data below show the number of people that live in each house in a small street.
4, 4, 5, 3, 2, 5, 11, 2
The outlier in this data set has:

A the greatest effect on the mean
B the greatest effect on the median
C the greatest effect on the mode
D an equal effect on the mean, median and mode

3 The two data sets below show the number of goals scored in 15 matches by two soccer teams.
Manchester: 0, 2, 1, 2, 1, 6, 0, 0, 1, 5, 0, 0, 1, 1, 1
Liverpool:



Which of the following statements is correct?

A The Manchester data are negatively skewed, while the Liverpool data are positively skewed.
B The Liverpool data are negatively skewed, while the Manchester data are positively skewed.
C Both sets of data are positively skewed.
D Both sets of data are negatively skewed.

4 The two-way table below shows the number of men and women who work in excess of 45 hours per week.

	Men	Women	Totals
≤ 45 hours	132	128	260
> 45 hours	69	34	103
Totals	201	162	**363**

The percentage of men who work greater than 45 hours per week is closest to:

A 28% **B** 34% **C** 51% **D** 67%

5 The figures below show the number of attempts that the boys and girls in a Year 12 class take to get their driver's licence.
Boys: 1, 2, 4, 1, 1, 2, 1, 1, 2, 2, 3, 1
Girls: 2, 2, 4, 2, 1, 2, 2, 3, 1, 1, 1, 2
When comparing the performance of the boys and the girls, it is found that the boys have:

A a lower mean and a lower standard deviation
B a lower mean and a higher standard deviation
C a higher mean and a lower standard deviation
D a higher mean and a higher standard deviation

SHORT ANSWER

1 Below are the ages of 15 players in a soccer squad.
23, 28, 25, 19, 17, 28, 29, 29, 22, 21, 35, 30, 22, 27, 26

a Calculate the mean age of the players in the squad.
b Find the median age of players in the squad.

2 The table below shows the number of house calls that a doctor has been required to make each day over a 32-day period.

Number of house calls	Frequency
0	1
1	6
2	8
3	9
4	6
5	2

a Copy the table into your workbook and add a cumulative frequency column.
b Calculate the mean number of house calls per day.

c Find the median number of house calls per day.
d What is the modal number of house calls per day?

3 The set of figures shown below shows the number of pages in a daily newspaper every day for two weeks.
72, 68, 76, 80, 64, 60, 132, 72, 84, 88, 60, 56, 76, 140
a What is the mean number of pages in the newspaper?
b What is the range?
c What is the interquartile range?
d Use the statistics function on your calculator to find the standard deviation.

4 The table below shows the number of rescues that are made each weekend at a major beach.

Number of rescues	Frequency
8	2
9	5
10	12
11	3
12	0
13	1
14	3

Use the statistics function on your calculator to find the mean and the standard deviation of these data.

5 The table below shows the customer waiting time at 10 am each morning at a bank over an 8-week period.

Waiting time	Class centre	Frequency	Cumulative frequency
0–1 minute		1	
1–2 minutes		4	
2–3 minutes		10	
3–4 minutes		13	
4–5 minutes		9	
5–6 minutes		3	

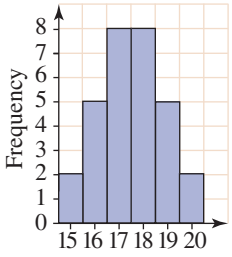
a Copy and complete the table.
b Use the statistics function on your calculator to find the mean and standard deviation.
c Draw a cumulative frequency histogram and polygon.
d Use the graph to estimate the interquartile range of the data.

6 The figures below show the number of points scored by a basketball player in six matches of a tournament.

36, 2, 38, 41, 27, 33

a Calculate the mean number of points per game.
b Calculate the median number of points per game.
c Explain why there is such a large difference between the mean and the median.

7 Consider the data set represented by the frequency histogram on the right.



a Are the data symmetrical?
b Can the mean and median of the data be seen?
c What is the mode of the data?

8 The table below shows the number of attempts that 20 members of a Year 12 class took to obtain a driver's licence.

Number of attempts	Frequency
1	11
2	4
3	2
4	2
5	0
6	1

a Show these data in a frequency histogram.
b Are the data positively or negatively skewed?

9 Draw an example of a frequency histogram for which the data are negatively skewed.

10 The figures below show the marks obtained by 20 students in English and Maths.
English: 56, 45, 57, 56, 65, 82, 74, 80, 91, 84, 68, 52, 67, 64, 60, 66, 74, 77, 77, 66
Maths: 65, 66, 58, 60, 61, 70, 74, 66, 69, 68, 71, 55, 51, 49, 50, 71, 99, 85, 70, 66
a Display the data in a back-to-back stem-and-leaf plot.
b For each subject find the median.
c For each subject state the range.
d For each subject find the interquartile range.

11 Betty runs a surf and ski shop. The table below shows the monthly sales of both types of equipment.

Month	Surf sales ($)	Ski sales ($)
January	20 000	5 000
February	18 000	6 000
March	12 000	8 000
April	9 000	10 000
May	6 000	12 000
June	4 000	12 000
July	5 000	9 000
August	8 000	8 000
September	10 000	6 000
October	11 000	3 000
November	15 000	4 000
December	22 000	9 000

a Display both sets of data on the same radar chart.
b Use the chart to compare trends in the sales.

12 The data below give the cost per minute of a long-distance telephone call with three companies.

	Telecomm	Omtus	Tel One
Day	21c	25c	17.5c
Economy	18c	15c	17.5c
Night	12c	12c	17.5c

Display this information in an area chart.

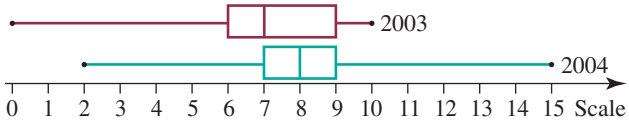
13 The stem-and-leaf plot above right compares the crowds (correct to the nearest thousand) at a football team's home and away matches.

Key: 2|5 = 25 000

Home		Away
8	**0**	6 7
7 3 2	**1**	0 1 1 6 8 9 9
6 6 3 2	**2**	4 5
5 5 2	**3**	

a Calculate the median of both data sets.
b Calculate the range of both data sets.
c Calculate the interquartile range of both data sets.
d Display both sets of data on a box-and-whisker plot.

14 The figure below shows a box-and-whisker plot showing the average number of weekly car sales made in 2003 and 2004.



a What was the median for each year?
b In which year was the range of sales greatest?
c In which year was the interquartile range of sales greatest?
d In which year did the car yard perform better? Explain your answer.

15 The two-way table below compares the number of men and women who are right- and left-handed.

	Men	Women	Totals
Right-handed	158	172	330
Left-handed	17	15	32
Totals	175	187	**362**

a What percentage of males are left-handed?
b What percentage of females are left-handed?
c Based on the above data, is there any significant difference between the percentage of male and female left-handers?

16 Hsiang compares her marks in 10 English exams and 10 Maths exams.
English: 76, 74, 80, 77, 73, 70, 75, 37, 72, 76
Maths: 80, 56, 92, 84, 65, 58, 55, 62, 70, 71
a Calculate Hsiang's mean mark in each subject.
b Calculate the range of marks in each subject.
c Calculate the standard deviation of marks in each subject.
d Based on the above data, in which subject would you say that Hsiang performs more consistently?

EXTENDED RESPONSE

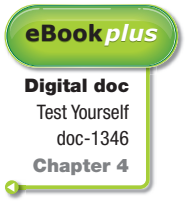
1 The data below show the weekly income among ten Year 12 boys and girls.
Boys: $80, $110, $75, $130, $90, $125, $100, $95, $115, $150
Girls: $50, $80, $75, $90, $90, $60, $250, $80, $100, $95

a Calculate the median of both sets of data.
b Calculate the range of both sets of data.
c Calculate the interquartile range of both sets of data.
d Display both sets of data on a box-and-whisker plot.
e Use the statistics function on the calculator to find the mean and standard deviation of both sets of data.
f Discuss whether the boys or girls have a more consistent average weekly income.

2 In the week leading up to the NRL grand final, Kylie records the number of points scored by both teams in each game throughout the season and displays the information on the stem-and-leaf plot below.
Key: 1|8 = 18

Sharks		Bulldogs
8	**0**	
8 4 4 2 2	**1**	5 5 5 8 8 8 9
8 8 6 4 4 4 3 2 2 0 0	**2**	0 0 2 2 2 2 6 6 6 8 8 8 9
8 8 6 2 0 0	**3**	0 0 0 2 2 2
8 6 2	**4**	

a Find the median of both sets of data.
b Which team's scores are the more consistent?
c Describe the skewness of the Sharks' scores.
d Find the mean and standard deviation of the Bulldogs' scores.



eBookplus ACTIVITIES

Are you ready?

Digital docs *(page 138)*
- SkillSHEET 4.1 (doc-1330): Finding the mean.
- SkillSHEET 4.2 (doc-1331): Finding the mode.
- SkillSHEET 4.3 (doc-1332): Finding the median.
- SkillSHEET 4.4 (doc-1336): Finding the range.
- SkillSHEET 4.5 (doc-1337): Finding the interquartile range.
- SkillSHEET 4.6 (doc-1338): Choosing the appropriate standard deviation.
- SkillSHEET 4.7 (doc-1341): Compiling a stem-and-leaf plot.
- SkillSHEET 4.9 (doc-1343): Drawing a box-and-whisker plot.

4A Measures of location and spread

Tutorial
- WE2 int-2417: Calculate the range, the interquartile range and the standard deviation. *(page 140)*

Digital docs
- SkillSHEET 4.1 (doc-1330): Finding the mean. *(page 143)*
- SkillSHEET 4.2 (doc-1331): Finding the mode. *(page 143)*
- SkillSHEET 4.3 (doc-1332): Finding the median. *(page 144)*
- Spreadsheet (doc-1333): One variable statistics. *(page 144)*
- GC program — Casio (doc-1334): UV Stats. *(page 144)*
- GC program — TI (doc-1335): UV Stats. *(page 144)*
- SkillSHEET 4.4 (doc-1336): Finding the range. *(page 144)*
- SkillSHEET 4.5 (doc-1337): Finding the interquartile range. *(page 144)*
- SkillSHEET 4.6 (doc-1338): Choosing the appropriate standard deviation. *(page 144)*
- Spreadsheet (doc-1339): Boxplots. *(page 144)*

4B Skewness

Tutorial
- WE5 int-2418: Determine measure of centre from a graph. *(page 148)*

Digital doc
- WorkSHEET 4.1 (doc-1340): Apply your knowledge of centre and spread to problems. *(page 151)*

4C Displaying multiple data sets

Tutorial
- WE10 int-2419: Learn how to display data in an information chart. *(page 155)*

Digital docs
- SkillSHEET 4.7 (doc-1341): Compiling a stem-and-leaf plot. *(page 156)*
- SkillSHEET 4.8 (doc-1342): Finding the mean, median, mode from a stem-and-leaf plot. *(page 156)*
- SkillSHEET 4.9 (doc-1343): Drawing a box-and-whisker plot. *(page 156)*
- Spreadsheet (doc-1344): Fast food sales. *(page 158)*

4D Comparison of data sets

Tutorial
- WE13 int-2420: Compare student performances on Maths tests. *(page 160)*

Digital docs
- WorkSHEET 4.2 (doc-1345): Apply your knowledge of statistics to problems. *(page 164)*

Chapter review

Digital doc
- Test Yourself (doc-1346): Take the end-of-chapter test to test your progress. *(page 169)*

To access eBookPLUS activities, log on to

www.jacplus.com.au